Volume 204, number 2 FEBS 3925 August 1986

Evolutionary conservation of the substrate-binding cleft of phosphoglycerate kinases

Nozomu Mori, Judith Singer-Sam and Arthur D. Riggs

Division of Biology, Molecular Biology Section, Beckman Research Institute of the City of Hope Duarte, CA 91010, USA

Received 13 May 1986

The primary structures of six phosphoglycerate kinases (PGKs) are known: three from mammals, one from yeast, and two from trypanosomes. Comparison of the amino acid sequence of these enzymes reveals 154 invariant positions out of 392 positions in the aligned sequences. Most of the conserved positions fall into the twelve β-sheets and adjacent peptide regions that form the inner loops surrounding the ATP and 3-phosphoglycerate-binding cleft. The homology between mammalian and yeast PGKs is greater than 94% for the inner-loop region, even though the overall homology is less than 65%. Trypanosome PGK has only 44% overall homology with the mammalian enzyme, but shows 74% homology in the inner-loop region. Trypanosome PGK contains a polypeptide segment in its N-terminal domain that is transposed in comparison with the other species.

Phosphoglycerate kinase *Nucleotide-binding protein* *Sequence homology* *Evolution*

1. INTRODUCTION

Phosphoglycerate kinase (PGK; ATP:3-phospho-D-glycerate 1-phosphotransferase, EC 2.7.2.3) catalyzes the reversible conversion of 1,3-diphosphoglycerate to 3-phosphoglycerate, with generation of one molecule of ATP [1]. This is a major enzyme in the glycolytic pathway and, therefore, is among the oldest housekeeping enzymes in nature. Considering the primary importance of PGK function for all living organisms, the structure of this enzyme is expected to be highly conserved during evolution.

X-ray crystallographic analysis of yeast [2,3] and horse muscle [4,5] enzymes has shown that they are very homologous in tertiary structure, even though homology at the amino acid level is only 65%. Human [6], yeast [7], and trypanosome [8] PGK protein sequences have recently been deduced from DNA sequences. We report here the amino acid sequence of mouse X-chromosome-linked PGK (PGK-1), and also compare the amino acid sequences of PGK from various sources to investigate which parts of the molecule are variable and conserved in relation to its secondary and tertiary structures. The substrate-binding cleft is found to be a very highly conserved region, and the molecular evolution of this enzyme is discussed.

2. MATERIALS AND METHODS

A full-length cDNA for PGK-1 was isolated from a mouse myeloma cDNA library by established procedures [9], using nick-translated cDNA for human PGK-1 [10] as a probe. Details of the cloning procedures and sequencing results are described elsewhere [18]. Amino acid sequences of human [6], horse [4], yeast [7], and trypanosome [8] PGKs were taken from the literature. Most parts of the sequence are well conserved in these species; therefore, alignment and placement of gaps were done by eye. The similarity of sequences was determined as percent identity calculated from the aligned regions only. The nomenclature and reference tertiary structure for PGK used was that of Banks et al. [4]. A slightly different structure

Published by Elsevier Science Publishers B.V. (Biomedical Division)

and nomenclature were described by Watson et al. [3] for yeast PGK.

3. RESULTS AND DISCUSSION

3.1. *Distribution of amino acid replacement*

Alignment of amino acid sequences of the six PGKs shows significant similarity in all sequences (fig.1). The homology between the three mammalian PGKs is greater than 96% over the entire 416 amino acid residues. Thus, PGK is a highly conserved protein. Yeast PGK, containing 415 amino acids, was aligned by introducing two gaps at positions 73 and 130 with respect to the mammalian PGK sequences. There are 148 amino acid changes in comparison with mouse PGK, i.e., 64% homology. The nonconserved residues are spread over the entire primary sequence (fig.1). However, when these positions were examined in view of the tertiary structure of the PGK molecule, it became quite clear that the distribution was uneven: in both the N- and C-domains, the inner-pocket region, which is composed of 12 β-sheets, 11 loops, and 6 helices, has undergone few amino acid replacements during evolution (fig.2).

PGK is a symmetrical two-domain protein with each domain mainly composed of six repeats of an α/β structural motif. The organization of these

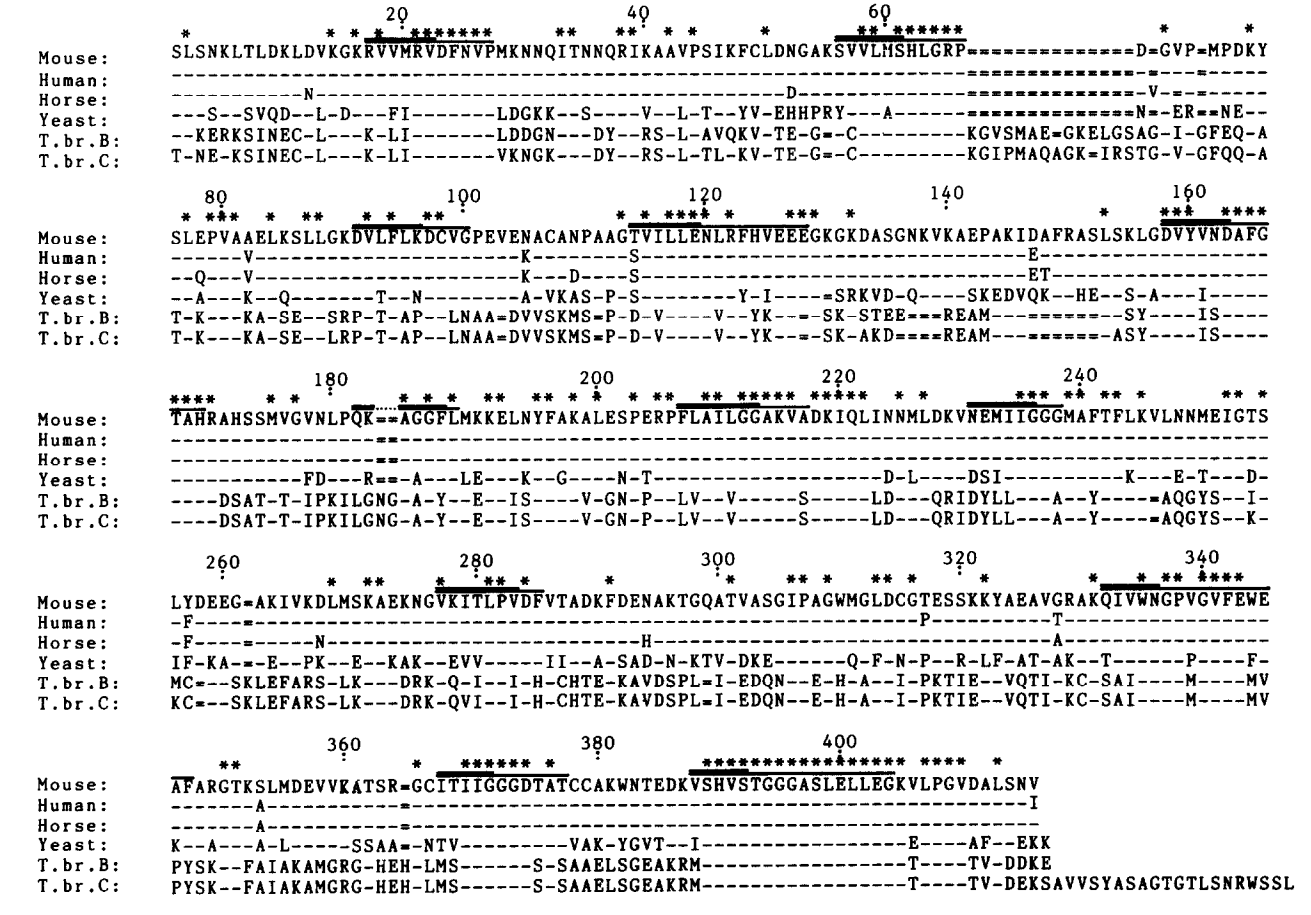


Fig.1. Sequence alignment for phosphoglycerate kinases. Extra gaps (=) are included to align yeast [7] and *Trypanosoma brucei* [8] PGK with mouse [18], human [6], and horse [4] PGK. The numbering refers to the mammalian PGKs. Amino acids are shown in the other sequences only when they differ from the mouse PGK. (–) Identical residues. The twelve β-sheets and adjacent peptides that form the inner loops in the substrate-binding cleft are indicated by double and single overlines, respectively, according to the structure of horse PGK [4]. (*) Invariant amino acid residues in the six PGKs.

motifs found in the C-domain follows the pattern that Rossmann pointed out for the NAD-binding site of lactate dehydrogenase; it is a common pattern for many dehydrogenases and kinases, and serves as a nucleotide-binding fold [11–13]. As expected from this β-sheet organization of PGK, the C-domain serves as the ATP-binding domain [3,4], whereas the 3-phosphoglycerate-binding site is located in the helix at position 393–401 (α14) at the bottom of the cleft [3]. Thus, the active site of this enzyme is believed to be in the C-domain. Considering the evolutionary conservation of amino acid residues, not only in the C-domain but also in the N-domain, some of the conserved residues in the inner loops and in a helix (α5) of the N-domain also might interact with substrates in the course of a dynamic reaction following the substrate-induced closing of the active site cleft as is postulated to occur in kinases [14].

3.2. *Transposition in trypanosome PGK*

Recently the sequences of two PGKs in trypanosome were reported [8]. At least 14 gaps had to be introduced to align the sequence with those from the other species. There are three major gaps, one between positions 66 and 67 in the mammalian sequence, the others between positions 136 and 152 in the trypanosome sequence (fig.1). The two portions, i.e., the 13-residue-long sequence KGVSMAEGKELGS in trypanosome PGK-B or KGIPMAQAGKIRS in trypanosome PGK-C and the 17-residue-long sequence, GNKVKAEPAKID AFRAS, in mouse PGK, as a representative of higher eukaryotic PGK, show no homology, but are located very close to each other in the outer edge of the N-domain of horse PGK (fig.2). It is reasonable to expect that trypanosome PGK should have a structure similar to that of horse and yeast PGKs because the size of the proteins is

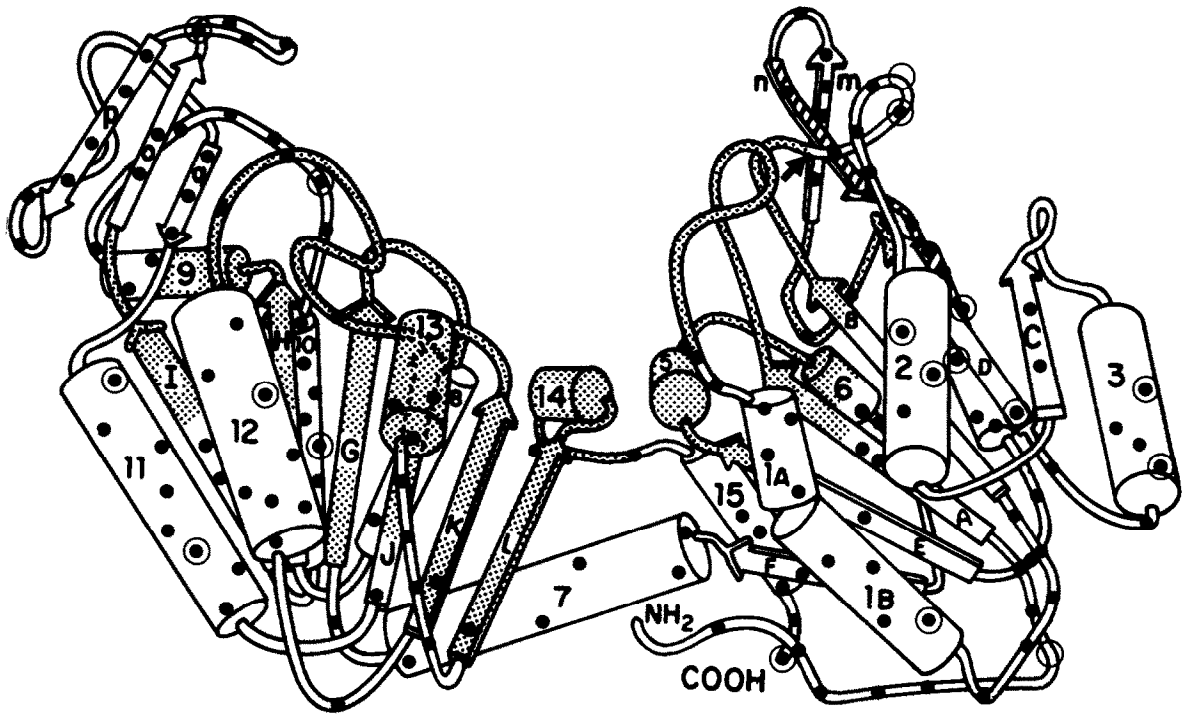


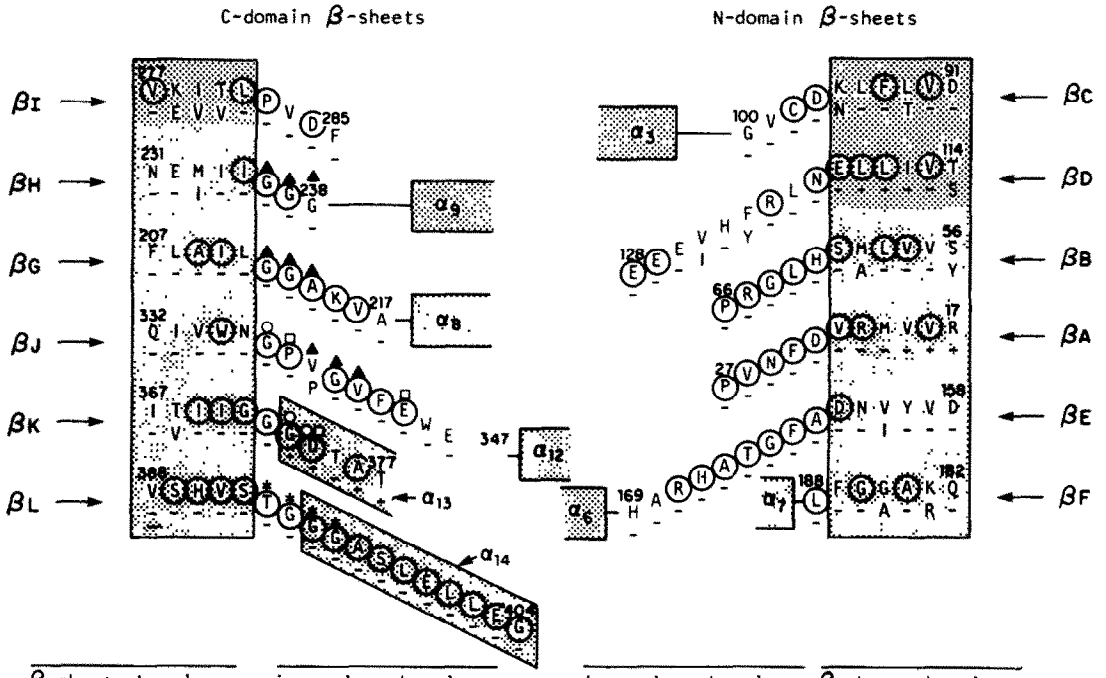
Fig.2. Distribution of amino acid substitutions in PGK. The positions of amino acid differences found among mammalian (mouse, human, and horse) PGKs are indicated by large open circles; those found between yeast PGK and one of the three mammalian PGKs are indicated by small filled circles. The three-dimensional drawing is derived from the structure for horse PGK by Banks et al. [4]. Their nomenclature for α-helical (α1–α15) and β-sheet (βA–βL, βm–βq) structures is shown. Dotted regions indicate conserved parts in which little substitution occurs. The arrow denotes the position where a large insertion occurs in trypanosome PGK, and hatched lines indicate the portion that includes two large deletions in trypanosome PGK.

similar, the amino acid homology is 44%, and if conservative substitutions are included, the homology increases to 65% [8]. In addition, secondary and tertiary structure usually is more strongly conserved than is primary structure [15]. Therefore, the additional 13-aa peptide found in trypanosome PGK in the outer-loop region between βn-sheet and $\alpha 2$-helix may be a structural compensation for the gap in a closely located segment including the βn-sheet and $\alpha 4$-helix (fig.2). Because their differing regions show no homology and are located in the outer surface of the domain, it seems unlikely that these residues contribute to substrate binding. It is more likely that the residues have potential structural importance in allowing correct folding of the molecule.

Disregarding this transposed portion and the gaps introduced in the aligned sequence, the comparison shows 154 invariant positions out of 392 positions of the six PGKs (fig.1).

3.3. *Evolutionary conservation of amino acids in the substrate-binding cleft*

The evolutionarily conserved region of PGK (see fig.2) is shown in more detail in fig.3. The conserved region consists of a cleft between the N- and C-domains, which is the substrate-binding pocket and is formed by two groups of six parallel β-pleated sheets (e.g., βA–βF and βG–βL) and adjacent peptides (e.g., inner-loop segments) (fig.3). In a total of 140 residues that fall into this portion of the three mammalian PGKs (residues 17–27,



	β-sheets homology:	inner loop homology:	inner loop homology:	β-sheets homology:
Yeast:	25/30=83.3%	40/41=97.6%	31/33=93.9%	28/36=77.8%
Trypanosome:	13/30=43.3%	31/41=75.6%	24/33=72.7%	18/36=50.0%

Fig.3. Homology in the substrate-binding cleft of PGK. The amino acid residues that form the six parallel β-sheets (βA–βF) and adjacent inner-loop segments in the N-domain of PGK are listed in the right part of the figure in accordance with topological order in the tertiary structure (see fig.2); those in the C-domain (βG–βL and adjacent segments) are on the left. For each region, the mammalian amino acid sequence is the upper sequence; yeast is the lower sequence, dashes indicate identity. The circled amino acid residues show conserved residues in all six species including trypanosome. The numbers indicate the positions of residues according to mammalian PGK. Symbols above the amino acid indicate substrate-binding residues: (▲) adenine-binding, (□) ribose-binding, (○) phosphate-binding, and (*) 3-phosphoglycerate-binding residues [3,4]. The six parallel β-sheet regions and α-helical regions are boxed and shaded.

56–66, 91–100, 114–128, 158–169, 182–188, 207–217, 231–238, 277–285, 332–347, 367–377, and 396–404), only one residue (114 at the outer edge of βD) is different between mouse and the others (Thr instead of Ser). The other 139 amino acids are identical among them. If the corresponding region of the yeast enzyme is compared with this region, it is evident that the inner-loop homology (the homology in the region of polypeptides that followed immediately after the series of six-parallel β-sheets in each domain) is 94% and 98%, and the β-sheet homology (the homology in the region of two series of the six parallel β-sheets) is 78% and 83% in the N-domain and in the C-domain, respectively, even though the overall homology between mammalian and yeast PGK sequences is less than 65%. Trypanosome PGK, which has only 44% overall homology with mammalian PGK, showed 73% and 76% homology in the inner-loop region in the N-domain and C-domains, respectively. It should be noted that the C-domain, which serves as an ATP-binding site, is more conserved than is the N-domain.

The essential conserved structure for nucleotide-binding activity in kinases and in NAD-dependent or NADP-dependent dehydrogenases is six parallel β-pleated sheets [11–13]. PGK contains two sets of six parallel β-pleated sheets, one in the N-domain and the other in the C-domain. However, only the topology of the latter set (βI-βH-βG-βJ-βK-βL) corresponds to that of the NADP-binding domain of dehydrogenases (βC-βB-βA-βD-βE-βF) [16,17]. Considering the difference in the topological order of these β-sheets, the six β-sheets set in the N-domain (βC-βD-βB-βA-βE-βF) are not thought to form an acceptable structure for an ATP-binding site. Rossmann and co-workers [11,13] predicted that the six parallel β-pleated sheets might be an essential feature in the evolution of kinases and dehydrogenases from an ancestral nucleotide-binding segment. In addition, our observations suggest that the inner-loop residues that follow immediately after each β-sheet structure are important for substrate binding or activity and were very conserved during the evolution of each of these nucleotide-binding enzymes.

ACKNOWLEDGEMENTS

This investigation was supported by a grant from the National Institutes of Health to A.D.R. (R01 GM31263). We thank Dr S. Ohno for discussion, and Shirley Pemberton for her help in preparing the manuscript.

REFERENCES

[1] Scopes, R.K. (1973) in: The Enzymes, 3rd edn, vol.8, pp.335–352, Academic Press, New York.
[2] Bryant, T.N., Watson, H.C. and Wendel, P.L. (1974) Nature 247, 14–17.
[3] Watson, H.C., Walker, N.P.C., Shaw, P.J., Bryant, T.N., Wendell, P.L., Fothergill, L.A., Perkins, R.F., Conroy, S.C., Dobson, M.J., Tuite, M.F., Kingsman, A.J. and Kingsman, S.M. (1982) EMBO J. 1, 1635–1640.
[4] Banks, R.D., Blake, C.C.F., Evans, P.R., Haser, R., Rice, D.W., Hardy, G.W., Merrett, M. and Phillips, A.W. (1979) Nature 279, 773–777.
[5] Rice, P.W. and Blake, C.C.F. (1984) J. Mol. Biol. 175, 218–223.
[6] Michelson, A.M., Markhamm, A.F. and Orkin, S.H. (1983) Proc. Natl. Acad. Sci. USA 80, 472–476.
[7] Hitzeman, R.A., Hagie, F.E., Hayflick, J.S., Chane, C.-Y., Seeburg, P.H. and Derynck, R. (1982) Nucleic Acids Res. 10, 7791–7808.
[8] Osinga, K.A., Swinkels, B.W., Gibson, W.C., Borst, P., Veeneman, G.H., Van Boom, J.H., Michels, P.A. and Opperdoes, F.R. (1985) EMBO J. 4, 3811–3817.
[9] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) in: Molecular Cloning: a Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[10] Singer-Sam, J., Simmer, R.L., Keith, D.H., Shively, L., Teplitz, M., Itakura, K., Gartler, S.M. and Riggs, A.D. (1983) Proc. Natl. Acad. Sci. USA 80, 802–806.
[11] Rossmann, M.G., Moras, D. and Olsen, K.W. (1974) Nature 250, 194–199.
[12] Rossmann, M.G., Liljas, A., Branden, C.-I. and Banaszak, L.J. (1975) in: The Enzymes, 3rd edn, vol.11, pp.61–102, Academic Press, New York.
[13] Rossman, M.G. and Argos, P. (1977) J. Mol. Biol. 109, 99–129.
[14] Anderson, C.M., Zucker, F.H. and Steitz, T.A. (1979) Science 200, 375–380.
[15] McLachlan, A.D. (1979) J. Mol. Biol. 128, 49–79.
[16] Levitt, M. and Chothia, C. (1976) Nature 261, 552–558.
[17] Richardson, J.S. (1977) Nature 268, 495–500.
[18] Mori, N., Singer-Sam, J., Lee, C.-Y. and Riggs, A.D. (1986) Gene, in press.